Screen Activity Study

Charlotte Xu

2024-02-05

Problem 1 a)

The purpose of collecting screen activity data here is to investigate the relationship between daily screen usage, social app usage and its potential implications on user's mental health. It is supported that Social media activity, such as time spent, was found to have a positive effect on the mental health domain. Karim, Fazida et al. "Social Media Use and Its Connection to Mental Health: A Systematic Review." Cureus vol. 12,6 e8627. 15 Jun. 2020, doi:10.7759/cureus.8627

Problem 1 b)

The Informed Consent Form is essential to ensure participants understand the study's purpose, risks, and benefits. It ensured the validation of the data source, the voluntary nature and the right to quit for participants.

Problem 1 c) Data Collection Plan:

Screen activity is recorded in real-time by my personal mobile device. I recorded daily entries of total screen time (Total.ST: total screen time in HH-MM format, and Total.ST.min: total screen time in MM format), social app usage screen time (Social.ST: social app screen time in HH-MM format, and Social.ST.min: social app screen time in MM format), total number of times the user picked up the phone (Pickups), the time of the first pick- up (Pickup.1st), and the total number of notification (Notification). For the data of first pickup, we use the record after user's wake-up to mark the beginning and end of the user's day. The collection period is from 01/14/2024-01/26/2024 (13 days), with the day of the week specified (Mon - Sun).

Problem 1 d)

```
ST <- readxl::read_excel("ScreenActivity_Data.xlsx")
ST$Date <- as.Date(ST$Date)
ST$Pickup.1st <- format(ST$Pickup.1st, format = "%H:%M")
ST$Daily_Proportion_Social_Screen_Time <- ST$Social.ST.min / ST$Total.ST.min
ST$Daily_Duration_Per_Use <- ST$Total.ST.min / ST$Pickups
head(ST)
```

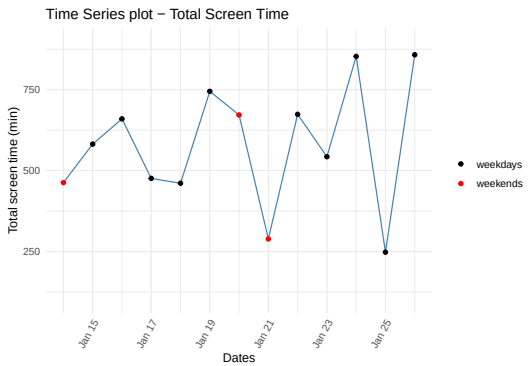
```
## # A tibble: 6 x 11
##   Date       Day   Total.ST Total.ST.min Social.ST Social.ST.min Pickups
##   <date>     <chr> <chr>           <dbl> <chr>             <dbl>   <dbl>
## 1 2024-01-14 Sun   7h43m             463 6h31m               391     124
## 2 2024-01-15 Mon   9h42m             582 8h24m               504      83
```

```
## 3 2024-01-16 Tue   11h               660 6h37m               397      60
## 4 2024-01-17 Wed   7h56m             476 5h8m                308     122
## 5 2024-01-18 Thu   7h41m             461 4h43m               283      97
## 6 2024-01-19 Fri   12h25m            745 10h48m              648     141
## # i 4 more variables: Pickup.1st <chr>, Notifications <dbl>,
## #   Daily_Proportion_Social_Screen_Time <dbl>, Daily_Duration_Per_Use <dbl>
```

Problem 2 a)

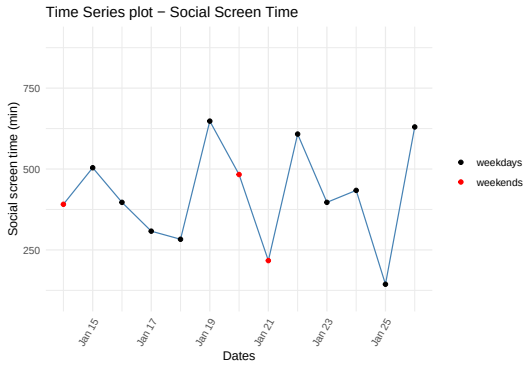
```
ST <- ST %>%
  mutate(weekend = ifelse(Day %in% c("Sun", "Sat"), TRUE, FALSE))

# Total screen time
total <-
  ggplot(ST, aes(
    x = Date,
    y = Total.ST.min,
    color = ifelse(weekend, "red", "black")
  )) +
  geom_line(color = "steelblue") +
  geom_point() +
  ylim(100, 900) +
  xlab("Dates") + ylab("Total screen time (min)") +
  labs(title = "Time Series plot - Total Screen Time") +
  scale_color_manual(labels = c("weekdays", "weekends"),
                     values = c("black", "red")) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 60, hjust = 1),
        legend.title = element_blank())
total
```

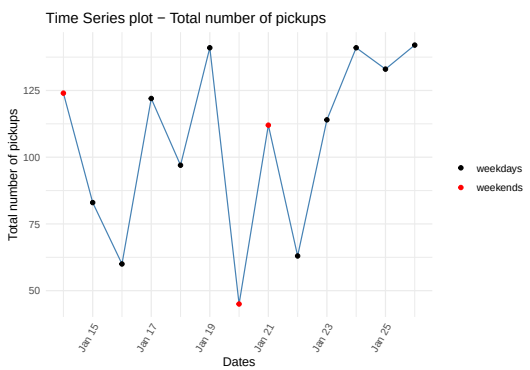



```
# Social screen time
social <-
  ggplot(ST, aes(
    x = Date,
    y = Social.ST.min,
    color = ifelse(weekend, "red", "black")
  )) +
  geom_line(color = "steelblue") +
```

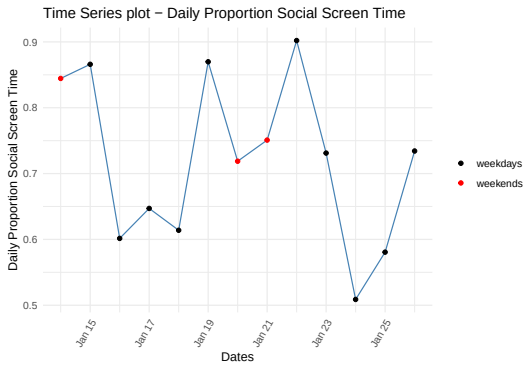
```
  geom_point() +
  xlab("Dates") + ylab("Social screen time (min)") +
  labs(title = "Time Series plot - Social Screen Time") +
  theme_minimal() +
  ylim(100, 900) +
  scale_color_manual(labels = c("weekdays", "weekends"),
                     values = c("black", "red")) +
  theme(axis.text.x = element_text(angle = 60, hjust = 1),
        legend.title = element_blank())
social
```



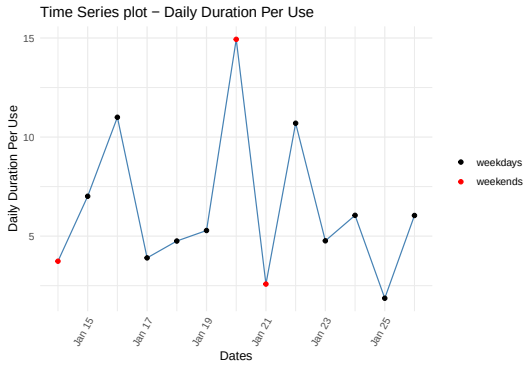

```
# Pickups
pickups <-
  ggplot(ST, aes(
    x = Date,
    y = Pickups,
    color = ifelse(weekend, "red", "black")
  )) +
  geom_line(color = "steelblue") +
  geom_point() +
  xlab("Dates") + ylab("Total number of pickups") +
  labs(title = "Time Series plot - Total number of pickups") +
  theme_minimal() +
  scale_color_manual(labels = c("weekdays", "weekends"),
                     values = c("black", "red")) +
  theme(axis.text.x = element_text(angle = 60, hjust = 1),
        legend.title = element_blank())
pickups
```



```r
# Daily_Proportion_Social_Screen_Time
PropST <-
  ggplot(ST,
         aes(
           x = Date,
           y = Daily_Proportion_Social_Screen_Time,
           color = ifelse(weekend, "red", "black")
         )) +
  geom_line(color = "steelblue") +
  geom_point() +
  xlab("Dates") + ylab("Daily Proportion Social Screen Time") +
  labs(title = "Time Series plot - Daily Proportion Social Screen Time") +
  theme_minimal() +
  scale_color_manual(labels = c("weekdays", "weekends"),
                     values = c("black", "red")) +
  theme(axis.text.x = element_text(angle = 60, hjust = 1),
        legend.title = element_blank())
PropST
```




```r
# Daily_Duration_Per_Use
Duration <-
  ggplot(ST, aes(
    x = Date,
    y = Daily_Duration_Per_Use,
    color = ifelse(weekend, "red", "black")
  )) +
  geom_line(color = "steelblue") +
  geom_point() +
  xlab("Dates") + ylab("Daily Duration Per Use") +
  labs(title = "Time Series plot - Daily Duration Per Use") +
  theme_minimal() +
  scale_color_manual(labels = c("weekdays", "weekends"),
                     values = c("black", "red")) +
  theme(axis.text.x = element_text(angle = 60, hjust = 1),
        legend.title = element_blank())
Duration
```



Total Screen Time

Overall the daily total screen time fluctuate around 550 minutes, with Jan.25 being the minimal point (250 minutes) and Jan.24 and Jan.26 being maximal points (around 800 minutes). It seems like if a high previous daily screen time is observed, the next following one would experience a big drop, and vice versa.

Social Screen Time

We can observe similar pattern with the total screen time usage. The social screen time also fluctuates, instead around 375mins. By Jan.19 the maximal point of 648 mins, and by Jan.25 the minimal point of 144 mins are observed. Worth noting that seems like the difference of weekends and weekdays did not impact the social screen time.

Pickups

The number of pickups seems fluctuated dramatically without a specific pattern, with a minimal point of 45 (Jan.20) and a maximal point of 142 (Jan.19). A dramatic increase in the number of pickups tends to be followed with a dramatic decrease in the next following day.

Daily Proportion of social screen time

For this social app usage ratio, seems like the mid-weekdays (Tuesday to Thursday) has a comparatively low value, indicating the density of the daily social usage is smaller. Mondays and Fridays tend to be observed having a comparatively higher proportional value, occupying values of 0.9 and 0.87.

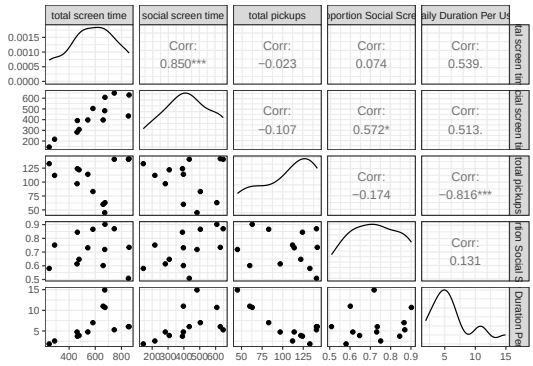
Daily Duration Per Use

From all the observations, the weekend of Jan.20 and Jan.21 tends to have the maximum and minimum value respectively, 15 minutes of screen use per pick up and 3 minutes of screen use per pick up. A high duration per use tends to experience a dramatic drop the next following day.

Problem 2 b)

```
ggpairs(
  ST,
  columns = c(
```

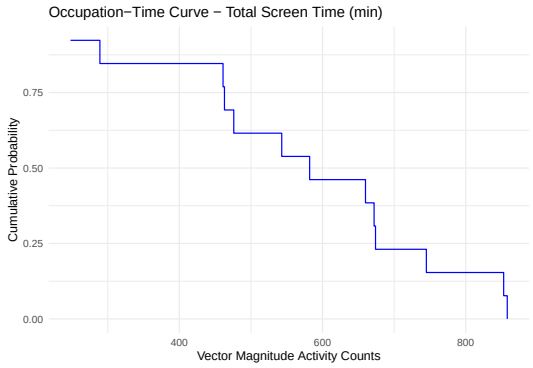
```
    "Total.ST.min",
    "Social.ST.min",
    "Pickups",
    "Daily_Proportion_Social_Screen_Time",
    "Daily_Duration_Per_Use"
  ),
  columnLabels = c(
    "total screen time",
    "social screen time",
    "total pickups",
    "Daily Proportion Social Screen Time",
    "Daily Duration Per Use"
  )
) +
  theme_bw()
```



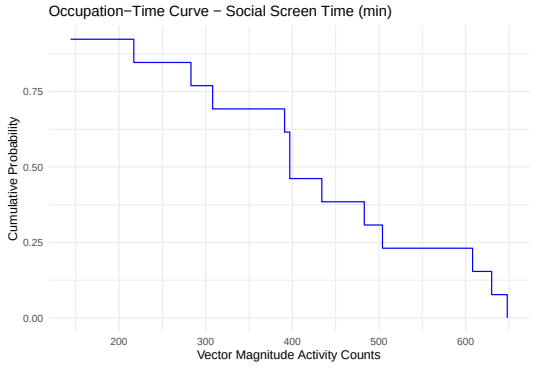
Total screen time and social screen time are most correlated with the score 0.85, as expected. Additionally, daily duration per use and total numbers of pickups displayed a strong negative correlation with the score -0.816. Except for total numbers of pickups has weak correlations with proportion social screen time, total screen time and social screen time, all other pairs displayed a moderate a correlation with the score ranged between 0.5 and 0.6.

Problem 2 c)

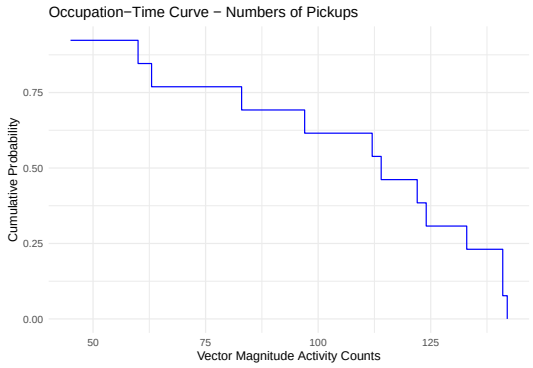
```
ecdf_values <- 1 - ecdf(ST$Total.ST.min)(ST$Total.ST.min)
ggplot(data.frame(x = ST$Total.ST.min, y = ecdf_values), aes(x, y)) +
  geom_step(color = "blue") +
  labs(
    title = paste("Occupation-Time Curve -", 'Total Screen Time (min)'),
    x = "Vector Magnitude Activity Counts",
    y = "Cumulative Probability"
  ) +
  theme_minimal()
```



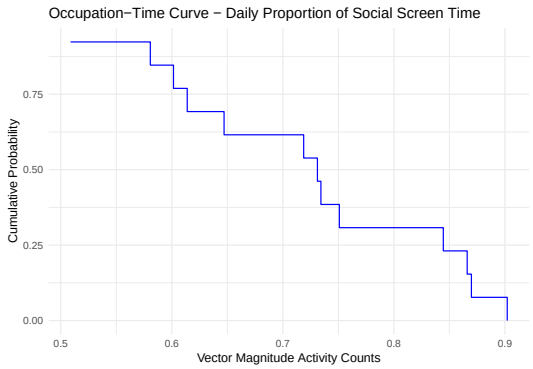

```
ecdf_values <- 1 - ecdf(ST$Social.ST.min)(ST$Social.ST.min)
ggplot(data.frame(x = ST$Social.ST.min, y = ecdf_values), aes(x, y)) +
  geom_step(color = "blue") +
  labs(
    title = paste("Occupation-Time Curve -", 'Social Screen Time (min)'),
    x = "Vector Magnitude Activity Counts",
    y = "Cumulative Probability"
  ) +
  theme_minimal()
```



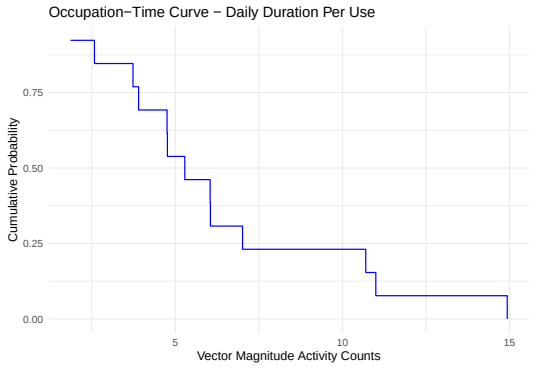

```
ecdf_values <- 1 - ecdf(ST$Pickups)(ST$Pickups)
ggplot(data.frame(x = ST$Pickups, y = ecdf_values), aes(x, y)) +
  geom_step(color = "blue") +
  labs(
    title = paste("Occupation-Time Curve -", 'Numbers of Pickups'),
    x = "Vector Magnitude Activity Counts",
    y = "Cumulative Probability"
  ) +
  theme_minimal()
```




```
ecdf_values <-
  1 - ecdf(ST$Daily_Proportion_Social_Screen_Time)(ST$Daily_Proportion_Social_Screen_Time)
ggplot(data.frame(
  x = ST$Daily_Proportion_Social_Screen_Time,
  y = ecdf_values
),
aes(x, y)) +
  geom_step(color = "blue") +
  labs(
    title = paste(
      "Occupation-Time Curve -",
      'Daily Proportion of Social Screen Time'
    ),
    x = "Vector Magnitude Activity Counts",
    y = "Cumulative Probability"
  ) +
  theme_minimal()
```



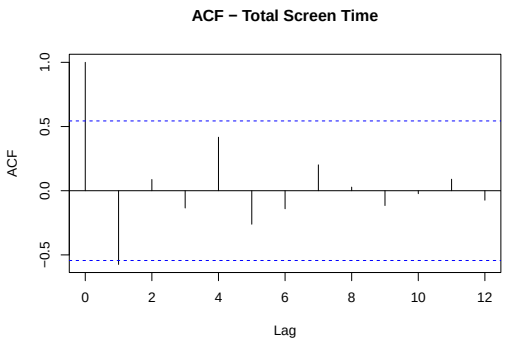

```
ecdf_values <-
  1 - ecdf(ST$Daily_Duration_Per_Use)(ST$Daily_Duration_Per_Use)
ggplot(data.frame(x = ST$Daily_Duration_Per_Use, y = ecdf_values),
       aes(x, y)) +
  geom_step(color = "blue") +
  labs(
    title = paste("Occupation-Time Curve -", 'Daily Duration Per Use'),
    x = "Vector Magnitude Activity Counts",
    y = "Cumulative Probability"
  ) +
  theme_minimal()
```



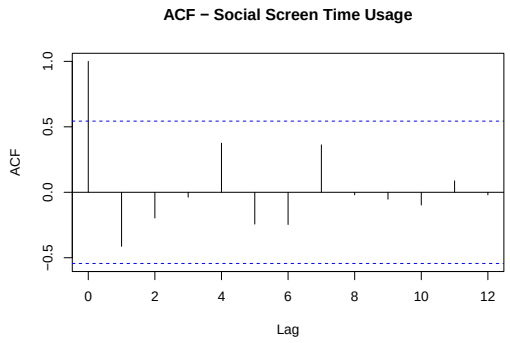
1. The Total Screen Time Occupation Time Curve starts with a plateau at 0.9, sharply decreasing around the median, suggesting infrequent occurrence of higher total screen time values.
2. The Social Screen Time Occupation Time Curve exhibits plateaus at 0.9, 0.7, and 0.25, forming three distinct usage modes. Overall the pattern created three major steps, indicating three mode of social screen time usage exist.
3. The Pickup Occupation Time Curve is generally flat, sharply rising after 0.6, indicating that most points occur within the initial 40% of the curve.
4. The Daily Proportion of Social Screen Time Occupation Time Curve closely mirrors the Social Screen Time Curve, with three plateaus at approximately 0.9, 0.6, and 0.3, indicating the value of daily proportion of social screen time mainly occurred at these three values.
5. The Daily Duration Per Use Occupation Time Curve sharply drops from 1 to 0.25 in the first 6 counts, with the subsequent range (6 to 15) representing only 0.25 of the total frequency, emphasizing rapid accumulation at smaller values.

Problem 2 d)

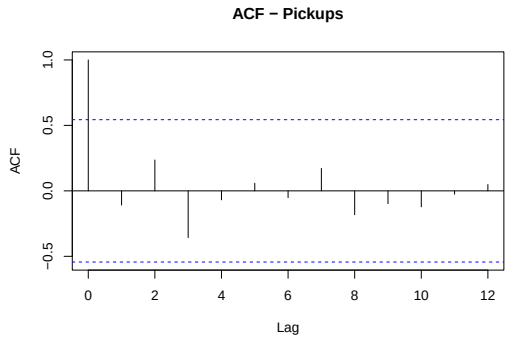
```
acf_tot <-
  acf(ST$Total.ST.min, main = "ACF - Total Screen Time", lag.max = 20)
```



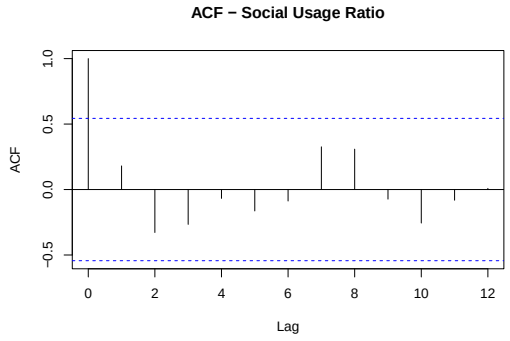

```
acf_social <-
  acf(ST$Social.ST.min, main = "ACF - Social Screen Time Usage", lag.max = 20)
```



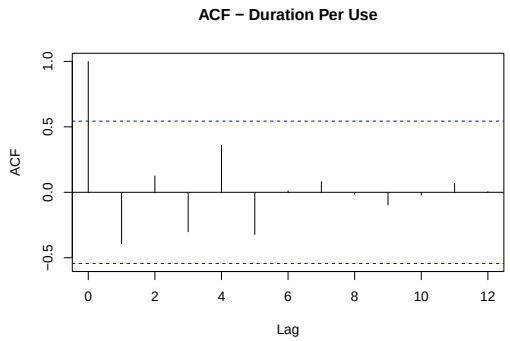

```
acf_pickups <- acf(ST$Pickups, main = "ACF - Pickups", lag.max = 20)
```




```
acf_Duration <-
  acf(ST$Daily_Proportion_Social_Screen_Time,
      main = "ACF - Social Usage Ratio",
      lag.max = 20)
```




```
acf_PropST <-
  acf(ST$Daily_Duration_Per_Use,
      main = "ACF - Duration Per Use",
      lag.max = 20)
```




```
acf(ST$Total.ST.min, plot = FALSE)
```

```
##
## Autocorrelations of series 'ST$Total.ST.min', by lag
##
##      0      1      2      3      4      5      6      7      8      9     10
##  1.000 -0.574  0.087 -0.135  0.417 -0.261 -0.141  0.201  0.028 -0.115 -0.023
##     11
##  0.090
```

```
acf(ST$Social.ST.min, plot = FALSE)
```

```
##
## Autocorrelations of series 'ST$Social.ST.min', by lag
##
##      0      1      2      3      4      5      6      7      8      9     10
##  1.000 -0.413 -0.196 -0.037  0.375 -0.243 -0.246  0.362 -0.019 -0.053 -0.097
##     11
##  0.087
```

```
acf(ST$Pickups, plot = FALSE)
```

```
##
## Autocorrelations of series 'ST$Pickups', by lag
##
##      0      1      2      3      4      5      6      7      8      9     10
##  1.000 -0.109  0.236 -0.357 -0.069  0.059 -0.052  0.172 -0.183 -0.098 -0.122
##     11
## -0.025
```

```
acf(ST$Daily_Proportion_Social_Screen_Time, plot = FALSE)
```

```
##
## Autocorrelations of series 'ST$Daily_Proportion_Social_Screen_Time', by lag
##
##      0      1      2      3      4      5      6      7      8      9     10
##  1.000  0.180 -0.328 -0.266 -0.068 -0.164 -0.088  0.326  0.308 -0.073 -0.256
##     11
## -0.082
```

```
acf(ST$Daily_Duration_Per_Use, plot = FALSE)
```

```
##
## Autocorrelations of series 'ST$Daily_Duration_Per_Use', by lag
##
##      0      1      2      3      4      5      6      7      8      9     10
##  1.000 -0.393  0.124 -0.302  0.359 -0.322  0.010  0.081 -0.014 -0.096 -0.021
##     11
##  0.069
```

High negative correlations at lag 1 can be observed in Total Screen time and social screen time with values -0.574 and -0.413 respectively, and being the peak point across all lags; Daily Duration per Use shows a relatively strong negative correlation (-0.393) at lag 1. Meanwhile, the number of pickups and proportion of social screen time exhibit stronger correlations at other lags (Lag 3 for pickups, lag 2, lag 7, and 8 for social use ratio). This implies potential exploration of longer-term patterns or cycles in these two datasets.

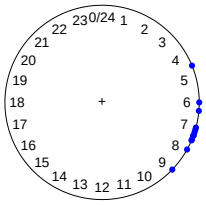
Problem 3 a)

```
ST$Pickup.1st <- as.POSIXct(ST$Pickup.1st, format = "%H:%M", tz = "UTC")
ST$Hour <- hour(ST$Pickup.1st)
ST$Minute <- minute(ST$Pickup.1st)
ST$Pickup_Angle <- circular((hour(ST$Pickup.1st)*60 + minute(ST$Pickup.1st)) / (24*60)*360)
ST$Pickup_Angle <- as.circular(ST$Pickup_Angle, type = "angles", units = "degrees")
head(ST)
```

```
## # A tibble: 6 x 15
##   Date       Day   Total.ST Total.ST.min Social.ST Social.ST.min Pickups
##   <date>     <chr> <chr>           <dbl> <chr>             <dbl>   <dbl>
## 1 2024-01-14 Sun   7h43m             463 6h31m               391     124
## 2 2024-01-15 Mon   9h42m             582 8h24m               504      83
## 3 2024-01-16 Tue   11h               660 6h37m               397      60
## 4 2024-01-17 Wed   7h56m             476 5h8m                308     122
## 5 2024-01-18 Thu   7h41m             461 4h43m               283      97
## 6 2024-01-19 Fri   12h25m            745 10h48m              648     141
## # i 8 more variables: Pickup.1st <dttm>, Notifications <dbl>,
## #   Daily_Proportion_Social_Screen_Time <dbl>, Daily_Duration_Per_Use <dbl>,
## #   weekend <lgl>, Hour <int>, Minute <int>, Pickup_Angle <circular>
```

Problem 3 b)

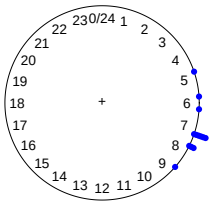
```
first.pickup.cir <- circular(ST$Pickup_Angle, units = "degrees",template = "clock24")
plot(first.pickup.cir, col = 'blue')
```



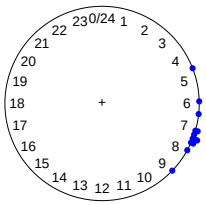
Most of the first pick up occurs between 6 to 8 am, with one occurence at 4:30 and one at 9:00.

Problem 3 c)

```
plot(first.pickup.cir, stack = TRUE, bins= 360/7.5, col='blue')
```




```
plot(first.pickup.cir, stack = TRUE, bins= 360/2.5, col='blue')
```



Increasing the bin size enhances stack division, providing a clearer data pattern. With a 30-minute bin size, the primary occurrence peak is observed at 7-7:30. In contrast, a 10-minute bin size plot reveals a more even spread of first pick-up occurrences between 7 to 7:50, leading to the conclusion that the majority of first pick-ups occurred during this time interval.

Problem 4 a)

By incorporating S_t into the distribution, we account for the fact that the number of pickups is expected to be proportional to the amount of time a person spends on the screen. This allows us to determine corresponding λ values for the Poisson model, where the expectation of $S_t\lambda$ may remain constant.

Problem 4 b)

```
poisson_model_b <- glm(Pickups ~ offset(log(Total.ST.min)), data = ST, family = poisson)
summary(poisson_model_b)
```

```
##
## Call:
## glm(formula = Pickups ~ offset(log(Total.ST.min)), family = poisson,
##     data = ST)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.69819    0.02695  -63.02   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 340.68  on 12  degrees of freedom
## Residual deviance: 340.68  on 12  degrees of freedom
## AIC: 426.45
##
## Number of Fisher Scoring iterations: 4
```

Problem 4 c)

```
ST$Xt <- ifelse(ST$Day %in% c("Mon", "Tue", "Wed", "Thu", "Fri"), 1, 0)
ST$Zt <- ifelse(as.Date(ST$Date) >= as.Date("2024-01-10"), 1, 0)

log_linear_model <- glm(Pickups ~  Xt + Zt + offset(log(Total.ST.min)), data = ST, family = poisson)
summary(log_linear_model)
```

```
##
## Call:
## glm(formula = Pickups ~ Xt + Zt + offset(log(Total.ST.min)),
##     family = poisson, data = ST)
##
## Coefficients: (1 not defined because of singularities)
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.62287    0.05965 -27.204   <2e-16 ***
## Xt          -0.09375    0.06687  -1.402    0.161
## Zt                NA         NA      NA       NA
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 340.68  on 12  degrees of freedom
## Residual deviance: 338.75  on 11  degrees of freedom
## AIC: 426.52
##
## Number of Fisher Scoring iterations: 5
```

c.1)

The p-value of 0.161 indicates a lack of statistical significance, suggesting no significant difference in the number of pickups between weekdays and weekends.

c.2)

The data is not valid for a p-value. (recorded starting from Jan.14, no valid data at Jan.10)

Problem 5 a)

```
vonmises_fit <- mle.vonmises(ST$Pickup_Angle)
vonmises_fit
```

```
##
## Call:
## mle.vonmises(x = ST$Pickup_Angle)
##
## mu: -0.1708  ( 2.536 )
##
## kappa: 0.1549  ( 0.394 )
```

Mu is estimated at -0.1708, suggesting a tendency for the circular data to cluster around this negative direction. A kappa value of 0.1549 indicates a relatively low concentration, implying that the data lacks specific clustering.

Problem 5 b)

```
probability_830_or_later <-
  1 - pvonmises(circular((8 * 60 + 30) / (24 * 60) * 360), mu = vonmises_fit$mu, kappa = vonmises_fit$ka
probability_830_or_later
```

```
## [1] 0.158642
```

The probability that your first pickup is 8:30AM or later is 0.158.